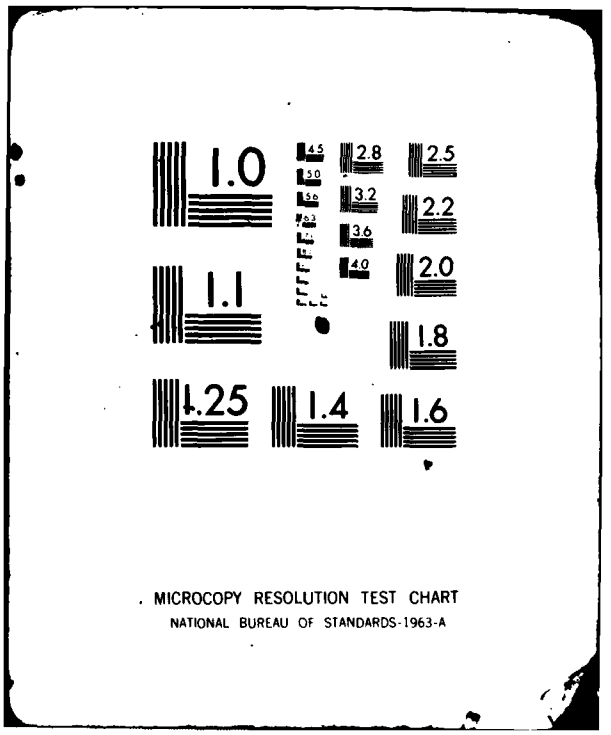
AD-A116 249 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1
THE CHOLESKY DECOMPOSITION ALGORITHM AND ARMA MODELING.(U)
APR 81 H J NEWTON, M PAGANO DAAG29-80-C-0070
UNCLASSIFIED TR-N-22 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843

INSTITUTE OF STATISTICS
Phone 713 - 845-3141

THE CHOLESKY DECOMPOSITION ALGORITHM AND ARMA MODELING

by H. Joseph Newton and Marcello Pagano

Institute of Statistics
Texas A&M University

Statistical Science Division
State University of New York
at Buffalo

Technical Report No. N-22

April 1981

Texas A & M Research Foundation
Project No. 4226T

"Robust Statistical Data Analysis and Modeling"

Sponsored by the Office of Naval Research

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER N-22	2. GOVT ACCESSION NO. AD-A116 249	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Cholesky Decomposition Algorithm and ARMA Modeling		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) H. Joseph Newton and Marcello Pagano		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0070 ONR N0001481MP10001
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas A&M University Institute of Statistics College Station, TX 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office P.O. Box 12277 Research Triangle Park, NC 27709		12. REPORT DATE April, 1981 13. NUMBER OF PAGES 10
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Autoregressive moving average model; prediction; likelihood estimation; Cholesky decomposition.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The history of the use of the Cholesky decomposition in the estimation and prediction of autoregressive moving average time series is reviewed and a modification is motivated by the results of Newton and Pagano (1981) in prediction theory and results in a significant reduction in the calculations required to evaluate the exact likelihood. It is also shown how the modified method can be used to obtain predictors and prediction variances.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The Cholesky Decomposition Algorithm and ARMA Modeling

By H. Joseph Newton and Marcello Pagano

Institute of Statistics, Texas A&M University and
Dept. of Biostatistics, Harvard University

Summary

The history of the use of the Cholesky decomposition in the estimation and prediction of autoregressive moving average time series is reviewed and a modification of Ansley's (1979) estimation method is suggested. This modification is motivated by the results of Newton and Pagano (1981) in prediction theory and results in a significant reduction in the calculations required to evaluate the exact likelihood. It is also shown how the modified method can be used to obtain predictors and prediction variances.

Some key words: Autoregressive moving average model; prediction; Likelihood estimation; Cholesky decomposition.

1. Introduction

Let $\{\varepsilon(t),\ t = 0, \pm 1, \ldots\}$ be a white noise time series of zero mean uncorrelated random variables having variance σ^2. Then the time series $\{Y(t),\ t = 0, \pm 1, \ldots\}$ satisfying

$$\sum_{j=0}^{p} \alpha(j)Y(t-j) = \sum_{k=0}^{q} \beta(k)\varepsilon(t-k),\ \alpha(0) = \beta(0) = 1\ ,$$

is called an autoregressive moving average process of order p, q (ARMA (p,q)). If $q = 0$, $Y(\cdot)$ is called an autoregressive process of order p, AR(p), while if $p = 0$ then $Y(\cdot)$ is a moving average process of order q, MA(q). Modern time series analysis has found the ARMA process extremely useful in modeling and predicting future values of a time series. However, the computational tasks involved in obtaining efficient estimators of the parameters $\alpha(\cdot)$, $\beta(\cdot)$, and σ^2 and in obtaining predictors of future values of $Y(\cdot)$ are extremely difficult and have led to the development of a wide variety of methods for accomplishing these tasks. These methods approach the problem from many points of view including 1) modified Newton Raphson likelihood maximization (Hannan (1969), Parzen (1971), Akaike (1973), for example), 2) nonlinear least squares (Box and Jenkins (1970), for example), 3) Kalman filtering techniques (Akaike (1974), Harvey and Phillips (1979), Jones (1980), Pearlman (1980), for example), 4) the use of the Cholesky matrix triangularization technique (Whittle (1963), Pagano and Parzen (1973), Pagano (1976), Pagano (1973), Phadke and Kedem (1978), Ansley (1979), Newton (1980), and Newton and Pagano (1981), for example), and others.

The purpose of this paper is to review the history of the Cholesky algorithm as applied to ARMA models and to show how one can easily modify the estimation procedure of Ansley (1979), which is currently regarded as the most computationally efficient exact likelihood estimation technique for at least the case where $p \geq q$ or $q < 5$ (see Pearlman (1980)), to obtain a significant reduction in the number of computations required.

Section 2 describes the ARMA process estimation and prediction problems and reviews the use of the Cholesky algorithm, while the proposed modification is given in section 3.

2. The Estimation and Prediction Problems

Given a realization $\underset{\sim}{Y}_T = (Y(1), \ldots, Y(T))^T$ from an ARMA(p,q) process, one often desires estimators of the $\alpha(\cdot)$, $\beta(\cdot)$, and σ^2 as well as memory-t, horizon-h, minimum mean square error linear predictors $Y(t + h|t)$ and prediction variances $\sigma^2_{t,h} = E\{Y(t+h) - Y(t+h|t)\}^2$ of $Y(t+h)$ given $\underset{\sim}{Y}_t = (Y(1),\ldots,Y(t))^T$ for a variety of values $h_1, \ldots, h_2$ and $t_1, \ldots, t_2$ of t and h.

The usual procedure is to obtain a Gaussian likelihood identification $\hat{\alpha}(\cdot)$, $\hat{\beta}(\cdot)$, and $\hat{\sigma}^2$ as the values of the parameters maximizing the Gaussian likelihood function

$$L(\underset{\sim}{\alpha}, \underset{\sim}{\beta}, \sigma^2 | \underset{\sim}{Y}_T) = (2\pi)^{-T/2} |\Gamma_{Y,T}|^{-1} \exp(-\underset{\sim}{Y}_T^T \Gamma_{Y,T}^{-1} \underset{\sim}{Y}_T)$$

where $\Gamma_{Y,t} = \mathrm{Cov}(\underset{\sim}{Y}_t) = (E(Y(j)Y(k))) = R_Y(|j-k|))$, $j, k = 1, \ldots, t$. The Gaussian assumption is often used also so that $Y(t+h|t)$ and $\sigma^2_{t,h}$, which in general are given by $Y(t+h|t) = \underset{\sim}{\lambda}_{t,h}^T \underset{\sim}{Y}_t$ and $\sigma^2_{t,h} = R_Y(0) - \underset{\sim}{r}_{t,h}^T \Gamma_{Y,t}^{-1} \underset{\sim}{r}_{t,h}$ where $\Gamma_{t,h}\, \underset{\sim}{\lambda}_{t,h} = \underset{\sim}{r}_{t,h} = (R_Y(t+h-1),\ldots,R_Y(h))^T$, can be obtained as the conditional mean and variance of $Y(t+h)$ given $\underset{\sim}{Y}_t$.

With the advent of efficient numerical algorithms for maximizing nonlinear functions that require a user to only supply a routine that evaluates the function for specified values of its arguments (see Dennis and More (1977), for example), the emphasis on finding the identification of $\alpha(\cdot)$, $\beta(\cdot)$, and σ^2 has been on finding efficient algorithms for evaluating L, or equivalently a more easily evaluated function having maxima at the same values of its parameters. The major efforts in this regard have been to apply the Kalman filter algorithm and the Cholesky decomposition algorithm. Each of these methods were originally used in ARMA modeling as a means of obtaining predictors and prediction variances.

2.1 The Cholesky Decomposition

Let A_n be a symmetric (n×n) matrix. Then A_n is positive definite if and only if it can be decomposed as the product $A_n = M_{A,n}M^T_{A,n}$ (Cholesky decomposition) or $A_n = L_{A,n}D_{A,n}L^T_{A,n}$ (modified Cholesky decomposition) where $M_{A,n}$ is an (n×n) lower triangular matrix with positive diagonal elements and $L_{A,n}$ is a unit lower triangular matrix while $D_{A,n}$ is a diagonal matrix containing the squares of the diagonal elements of $M_{A,n}$. Further these decompositions are unique and nested, i.e., for example, $L_{A,K}$ is the lower triangular matrix in the decomposition of the K×K principal minor of A_n, $K = 1, \ldots, n$. Thus the $(j,k)^{th}$ element of $L_{A,K}$ can be denoted $L_{A,j,k}$ for $j, k \leq K$.

2.2 The Cholesky Decomposition and Prediction

In his seminal work on time series prediction, Whittle (1963) suggested that such a triangularization would be useful in finding $Y(t+h|t)$ and $\sigma^2_{t,h}$. Pagano and Parzen (1973) and Pagano (1976) showed that for an MA(q) process

$$Y(t+h|t) = \begin{cases} \sum_{k=h}^{q} L_{Y,t+h,t+h-k}\, e(t+h-k), & h=0, \ldots, q \\ 0, & h>q \end{cases}$$

$$\sigma^2_{t,h} = \sum_{k=0}^{h-1} L^2_{Y,t+h,t+h-k}\, D_{Y,t+h-k,t+h-k}$$

$$L_{Y,K,K-j} = 0 \quad , j>q$$

$$\lim_{K\to\infty} L_{Y,K,K-j} = \beta(j) \quad , j = 1, \ldots, q,$$

$$\lim_{K\to\infty} D_{Y,K,K} = \sigma^2,$$

where $e(1) = Y(1)$ and $e(j) = Y(j) - \sum_{k=1}^{\min(j-1,q)} L_{Y,j,j-k}\, e(j-k)$, $j>2$ is the jth element of $L^{-1}_{Y,K}\, \underset{\sim}{Y}_K$ for any $K \geq j$. *Thus one need only calculate the* successive rows of L_Y and the corresponding $e(\cdot)$ until convergence of its $q + 1$ nonzero elements to the $\beta(\cdot)$ is achieved, at which point one replaces $L_{Y,K,K-j}$ and $D_{Y,K,K}$ by $\beta(j)$ and σ^2 respectively in the above formulas. Newton (1980) discusses the extension of this algorighm to the multivariate MA(q) process.

Pagano and Parzen (1973) further suggested that this procedure could be extended to the ARMA (p,q) case by applying the autoregressive filter $\alpha(\cdot)$ to Y(p+1), ..., Y(T) and using the MA(q) algorithm combined with an AR(p) prediction algorithm.

Finally Newton and Pagano (1981) show that for a general purely nondeterministic covariance stationary time series Y(·)

$$Y(t+h|t) = \sum_{k=h}^{t+h-1} L_{Y,t+h,t+h-k}\, e(t+h-k)$$

$$\sigma^2_{t,h} = \sum_{k=0}^{h-1} L^2_{Y,t+h,t+h-k}\, D_{Y,t+h-k,t+h-k}$$

$$\lim_{K\to\infty} L_{Y,K,K-j} = \beta_\infty(j) \quad , \quad j \geq 0$$

$$\lim_{K\to\infty} D_{Y,K,K} = \sigma^2_\infty \quad ,$$

where the $\beta_\infty(\cdot)$ and σ^2_∞ are the coefficients and the noise variance of the MA(∞) representation of $Y(\cdot)$ and $e(1) = Y(1)$, while

$$e(j) = Y(j) - \sum_{k=1}^{j-1} L_{Y,j,k}\, e(k) \quad , \quad j \geq 2 \ .$$

Applying this general theory to the ARMA (p,q) process and further factoring L_Y into $L_Y = L_Z L_X$ where $Z(\cdot)$ is an AR(p) process with coefficients $\alpha(\cdot)$ and noise variance σ^2 and $\Gamma_{X,K} = L^{-1}_{Z,K}\, \Gamma_{Y,K} L^{-T}_{Z,K} = L_{X,K} D_{X,K} L^T_{X,K}$ has last K-q rows and columns being the covariance matrix of an MA(q) process, Newton and Pagano obtain

$$Y(t+h|t) = X(t+h|t) - \sum_{j=1}^{p} \alpha(j) Y(t+h-j|t) \ ,$$

$$\sigma^2_{t,h} = \sum_{k=0}^{h-1} D_{X,t+h-k,t+h-k} \left\{ \sum_{\ell=t+h-k}^{t+h} L_{Z,t+h,\ell} L_{X,\ell,t+h-k} \right\}^2 \quad ,$$

$$Y(t+h-j|t) = Y(t+h-j) \text{ if } j \geq h \ ,$$

$$X(t+h|t) = \begin{cases} \sum_{k=h}^{q} L_{X,t+h,t+h-k}\, e(t+h-k), & h = 1, \ldots, q \\ 0, & h > q \end{cases} \ ,$$

$e(j)$ is the j^{th} element of $L_{X,K}^{-1} X_K$, $K \geq j$, and $X_K = L_{Z,K}^{-1} Y_K$. Further,

$$\lim_{K\to\infty} L_{X,K,K-j} = \beta(j) \quad , \quad j = 1, \ldots, q \ ,$$

$$\lim_{K\to\infty} D_{X,K,K} = \sigma^2 \ .$$

A number of other results are given by Newton and Pagano which greatly reduce the amount of storage and computations necessary for obtaining L_X and L_Z . We note that this method is different than that given by Pagano and Parzen (1973) in that instead of leaving the initial values $Y(1), \ldots, Y(p)$ untransformed, the Newton and Pagano algorithm forms the time series $X(\cdot)$ by $X_K = L_{Z,K}^{-1} Y_K$, i.e.

$$X(j) = \begin{cases} Y(j) & , j = 1 \\ \sum_{k=0}^{j-1} \alpha_{j-1}(k) Y(j-k) & , j = 2, \ldots, p \\ \sum_{k=0}^{p} \alpha(k)\, Y(j-k) & , j > p \end{cases}$$

where the $\alpha_j(k)$ are easily obtained by performing the Levinson (1974) recursion for $j = p-1, \ldots, 1$, with $\alpha_{p+1}(k) = \alpha(k)$:

$$\alpha_j(i) = \frac{\alpha_{j+1}(i) - \alpha_{j+1}(j+1)\alpha_{j+1}(j+1-i)}{1 - \alpha_{j+1}^2(j+1)} \quad , \ i = 1, \ldots, j < p$$

2.3 The Cholesky Decomposition and Parameter Estimation

Pagano (1973) discusses the use of the Cholesky decomposition in solving the Yule Walker equations for an AR(p) process. Phadke and Kedem (1978) and Ansley (1979) apply the Cholesky decomposition to the problem

of evaluating the likelihood function of the MA(q) and ARMA(p,q) processes respectively, essentially using the same transformation and decomposition that Pagano and Parzen (1973) used for obtaining predictors. We briefly describe Ansley's method and show in the next section how it can be improved by incorporating the theory of section 2.2.

Define a new time series, with $m = \max(p, q+1)$,

$$W(t) = \begin{cases} Y(t) \ , & t = 1, \ldots, m \\ \sum_{j=0}^{p} \alpha(j) Y(t-j) \ , & t = m+1, \ldots, T \ . \end{cases}$$

Thus in matrix notation this is $\underset{\sim}{W}_T = M_{Y,T} \underset{\sim}{Y}_T$ where $M_{Y,T}$ is a (TXT) unit lower triangular matrix and $\Gamma_{W,T} = \mathrm{Cov}(\underset{\sim}{W}_T) = M_{Y,T} \Gamma_{Y,T} M_{Y,T}^T$, which is a banded matrix of bandwidth q, i.e., there are q nonzero subdiagonals, from row m+1 on. Thus finding the Cholesky decomposition $\Gamma_{W,T} = M_{W,T} M_{W,T}^T$ and the new random vector $\underset{\sim}{e}_T = M_{W,T}^{-1} \underset{\sim}{W}_T$ can be done recursively only having to compute q elements in each row of $M_{W,T}$. Then $\underset{\sim}{e}_T = M_{W,T}^{-1} M_{Y,T} \underset{\sim}{Y}_T$ and thus

$$\underset{\sim}{e}_T \sim N_T(\underset{\sim}{0}, \ M_{W,T}^{-1} \ M_{Y,T} \Gamma_{Y,T} M_{Y,T}^T M_{W,T}^{-T} = I_T) \ .$$

Thus maximizing L is equivalent to maximizing

$$\rho(\underset{\sim}{\alpha}, \ \underset{\sim}{\beta}, \ \sigma^2 | \underset{\sim}{e}_T) = (2\pi)^{-T/2} |M_{W,T}|^{-1} \exp\left(- \sum_{t=1}^{T} \frac{e^2(t)}{2\sigma^2}\right) \ .$$

3. Using the Prediction Results in Ansley's Method

Thus it is clear that Ansley's method consists of factoring the lower triangular matrix in the Cholesky decomposition of $\Gamma_{Y,T}$ into the product of the two easily calculated lower triangular matrices $M_{Y,T}$ and $M_{W,T}$. But from section 2.2 it is easily seen that if this factorization is done as

$L_{Y,T} = L_{Z,T}L_{X,T}$ one can call upon the results of Newton and Pagano to halt calculation of successive rows of $L_{X,T}$ once their elements have converged to $\beta(1), \ldots, \beta(q)$. The experience of the present authors is that the speed of this convergence, which is a function of how close the zeros of $\sum_{k=0}^{q} \beta(k)z^k$ are to the unit circle, is such that a significant reduction in computations is obtained, particularly if T is large relative to q. Further one can then use the prediction theory of section 2.2 to obtain desired predictors and prediction variances.

We note that section 2.2 can also be used to prove that the elements of the rows of Ansley's $M_{W,T}$ also converge to the $\beta(\cdot)$ and our experience is that the rate of convergence is virtually identical with the above method.

Thus either factorization could be used but it seems more natural to factor $L_{Y,T}$ into an exact AR(p) covariance matrix factor $L_{Z,T}$ and an exact MA(q) (except for the first max(p,q+1) rows and columns) covariance matrix factor $L_{X,T}$ and use the methods of section 2.2 to find the likelihood, the predictors, and the prediction variances.

REFERENCES

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models, Biometrika, 60, 255-265.

Akaike, H. (1974). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. Annals of the Institute of Statistical Mathematics, 26, 363-387.

Ansley, C. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. Biometrika, 66, 59-65.

Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis forecasting and control. Holden-Day, San Francisco.

Dennis, J. E. and More, J. J. (1977). Quasi-Newton methods, motivation and theory. SIAM Rev., 19, 46-89.

Hannan, E. J. (1969). The estimation of mixed moving average autoregressive systems. Biometrika, 56, 579-593.

Harvey, A. C. and Phillips, G. D. (1979). Maximum likelihood estimation of regression models with autoregressive - moving average disturbances. Biometrika, 66, 49-58.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. Technometrics, 22, 389-395.

Levinson, N. (1974). The Wiener RMS error criterion in filter design and prediction. Journal of Mathematical Physics, 15, 261-278.

Newton, H. J. (1980). Efficient estimation of multivariate moving average autocovariances. Biometrika, 67, 227-231.

Newton, H. J. and Pagano, M. (1981). The finite memory prediction of covariance stationary time series. Submitted for publication.

Pagano, M. (1973). An algorithm for fitting autoregressive schemes. Applied Statistics, 21, 274-281.

Pagano, M. (1976). On the linear convergence of a covariance factorization algorithm. Journ. Assoc. Comp. Mach., 23, 310-316.

Pagano, M. and Parzen, E. (1973). Timesboard; A time series package. Proc. of Comp. Sci. and Stat.: 7th Ann. Symp. on the Interface, ed. by W. J. Kennedy, Statistical Laboratory, Iowa State Univ., Ames, Iowa.

Parzen, E. (1971). Efficient estimation of stationary time series mixed schemes. Proc. 39th Session of I.S.I., Washington, D.C..

Pearlman, J. G. (1980). An algorithm for the exact likelihood of high-order autoregressive - moving average process. Biometrika, 67, 232-233.

Phadke, M. S. and Kedem, G. (1978). Computation of the exact likelihood function of multivariate moving average models. Biometrika, 65, 511-519.

Whittle, P. (1963). Prediction and Regulation by linear least squares methods. English Universities Press, London.


FILMED
8